A. Examine window of typed SNPs

Reference haplotypes aligned with genotypes at the typed SNPs (untyped SNPs not needed)

Reference haplotypes

h_1	1	1	0	0	1	0
h_2	1	1	0	0	1	0
h_3	1	1	0	0	1	0
h_4	1	0	1	0	0	1
h_5	1	0	1	0	0	1
h_6	1	0	1	0	0	1
h_7	0	0	1	0	1	0
h_8	0	0	1	0	1	0

Sample's genotypes are unphased. One marker is missing due to GWAS error

Target genotype

x_1	2	1	0	0	?	1

B. Find optimal haplotype pairs in each window

Unique reference haplotypes

Find unique haplotypes

h_1	1	1	0	0	1	0
h_4	1	0	1	0	0	1
h_7	0	0	1	0	1	0

Find h_1, h_4 using least squares

$$h_1 + h_4 \approx x_1$$

Extract matching haplotypes

$$\|x_1' - h_1 - h_4\|^2 = 2$$
$$\|x_1' - h_1 - h_7\|^2 = 2$$
$$\|x_1' - h_4 - h_7\|^2 = 3$$

Initialize missing

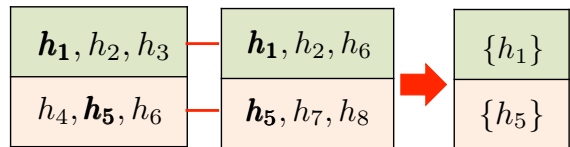|						
x_1'	2	1	0	0	0	1

strand 1: $\{h_1, h_2, h_3\}$
strand 2: $\{h_4, h_5, h_6\}$

Candidate haplotypes

C. Connect neighbors in 1 of 2 ways

Parallel connection generates 2 surviving haplotypes:

$\boldsymbol{h_1}, h_2, h_3$	—	$\boldsymbol{h_1}, h_2, h_6$	➡	$\{h_1\}$
$h_4, \boldsymbol{h_5}, h_6$	—	$\boldsymbol{h_5}, h_7, h_8$		$\{h_5\}$

A switch at window 2 generates 1 surviving haplotype:

h_1, h_2, h_3	✕	$h_1, h_2, \boldsymbol{h_6}$	➡	$\{\}$
$h_4, h_5, \boldsymbol{h_6}$		h_5, h_7, h_8		$\{h_6\}$

Window 1   **Window 2**   **Survivors**
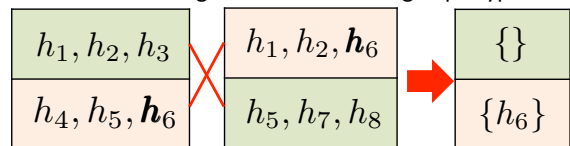
D. Stitch window-by-window from left to right

Unphased haplotypes

$\boldsymbol{h_1}, h_2, h_3$	$\boldsymbol{h_1}, h_2, h_6$	$\boldsymbol{h_1}, h_3$	$h_4, \boldsymbol{h_5}, h_8$
$h_4, \boldsymbol{h_5}, h_6$	$\boldsymbol{h_5}, h_7, h_8$	$h_2, \boldsymbol{h_5}$	h_2, h_6

Switch

Phased haplotypes

Breakpoint

$\boldsymbol{h_1}$	$\boldsymbol{h_1}$	$\boldsymbol{h_1}$	h_2, h_6
$\boldsymbol{h_5}$	$\boldsymbol{h_5}$	$\boldsymbol{h_5}$	$\boldsymbol{h_5}$

Window 1   **Window 2**   **Window 3**   **Window 4**